

External validation of risk prediction models incorporating common genetic variants for incident colorectal cancer using UK Biobank

Catherine L Saunders¹, Britt Kilian², Deborah J Thompson³, Luke J McGeoch⁴, Simon J Griffin⁵, Antonis C Antoniou⁶, Jon D. Emery⁷, Fiona M Walter⁸, Joe Dennis⁹, Xin Yang¹⁰, Juliet A Usher-Smith¹¹

¹ Senior Research Associate, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

² Research Associate, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

³ Principal Research Associate, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Wort's Causeway, Cambridge, CB1 8RN, UK

⁴ Medical Student, School of Clinical Medicine, University of Cambridge, Cambridge, UK

⁵ Professor of General Practice, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

⁶ Professor of Cancer Risk Prediction, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Wort's Causeway, Cambridge, CB1 8RN, UK

⁷ Professor of Primary Care Cancer Research, Department of General Practice, Centre for Cancer Research, Faculty of Medicine, Dentistry & Health Sciences The University of Melbourne, Victorian Comprehensive Cancer Centre, Melbourne Victoria, 3010, Australia

⁸ Principal Researcher in Primary Care Cancer Research, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, 2 Wort's Causeway, Cambridge, CB1 8RN, UK

⁹ Research Associate, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Wort's Causeway, Cambridge, CB1 8RN, UK

¹⁰ Research Associate, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Wort's Causeway, Cambridge, CB1 8RN, UK

¹¹ Senior Clinical Research Associate, The Primary Care Unit, Institute of Public Health, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

Running title: Genetic risk prediction models for colorectal cancer

Key words: Colorectal cancer, risk, prediction, genetics, external validation

Financial support

This research has been conducted using the UK Biobank Resource under Application Number 28126. This work was funded by a grant from Bowel Cancer UK (18PG0008). J Usher-Smith is funded by a Cancer Research UK Prevention Fellowship (C55650/A21464). ACA and XY are supported by a Cancer Research UK grant (C12292/A20861). DE was supported by a Cancer Research UK grant (C1287/A16563). The University of Cambridge has received salary support in respect of SJG from the NHS in the East of England through the Clinical Academic Reserve. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. All researchers were independent of the funding body and the funder had no role in data collection, analysis and interpretation of data; in the writing of the report; or decision to submit the article for publication.

Correspondence to:

Dr Juliet Usher-Smith, jau20@medschl.cam.ac.uk

The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Box 113 Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK

Competing Interests:

JDE has received a research grant from Genotype Pty Ltd for research on the implementation of polygenic risk scores for colorectal cancer. Since completing her contribution to this paper

DJT has left Cambridge University and is now employed by Genomics plc. The other authors have nothing to declare.

Word count: 4315

Figures and Tables: 6

ABSTRACT

The aim of this study was to compare and externally validate risk scores developed to predict incident colorectal cancer (CRC) that include common genetic variants (single nucleotide polymorphisms, SNPs), with or without established lifestyle/environmental (questionnaire-based/classical/phenotypic) risk factors. We externally validated 23 risk models from a previous systematic review in 443,888 participants aged 37-73 from the UK Biobank cohort who had 6 year prospective follow-up, no prior history of CRC and data for incidence of CRC through linkage to national cancer registries. There were 2,679 (0.6%) cases of incident CRC. We assessed model discrimination using the Area Under the Curve (AUC) and relative risk calibration. The AUC of models including only SNPs increased with the number of included SNPs and was similar in men and women: the model by Huyghe with 120 SNPs had the highest AUC of 0.62(95%CI 0.59-0.64) in women and 0.64(95%CI 0.61-0.66) in men. Adding phenotypic risk factors without age improved discrimination in men but not in women. Adding phenotypic risk factors and age increased discrimination in all cases ($p<0.05$), with the best performing models including SNPs, phenotypic risk factors and age having AUCs between 0.64-0.67 in women and 0.67-0.71 in men. Relative risk calibration varied substantially across the models. Among middle-aged people in the UK, existing polygenic risk scores discriminate moderately well between those who do and do not develop colorectal cancer over six years. Consideration should be given to exploring the feasibility of incorporating genetic and lifestyle/environmental information in any future stratified CRC screening program.

INTRODUCTION

Colorectal cancer (CRC) is one of the most commonly diagnosed malignancies worldwide. Incidence is rising in many countries[1], particularly among younger individuals[2]. There is good evidence that population screening reduces CRC incidence and mortality[3–6], and in most countries with a high frequency of CRC screening has now been introduced[7].

As for other cancers, including breast[8], lung[9] and ovarian[10], there is increasing interest in using risk models to predict who is at highest and lowest risk of developing CRC and guide screening decisions. By identifying those who are more likely to benefit from screening and inviting them earlier, more frequently or for different screening tests, this approach has the potential to increase net benefit of CRC screening at lower cost.

We, and others, have previously shown that several risk models based on phenotypic variables have relatively good discrimination in external validation and could, therefore, be used for this purpose[11,12]. Advances in genetic research and technology mean that it will soon be possible to provide a relatively cheap, quick and accurate assessment of an individual's genetic risk of CRC. Models incorporating genetic variables may, therefore, be simpler to implement. Previous work has also shown that using such genetic information to stratify screening has the potential to improve efficiency of screening[13] and reduce the number of individuals screened while still detecting as many cases[14]. Of the 29 models identified in our review, however, many had only been assessed in the development population. The comparative performances of these models in the same population and the added value of including phenotypic risk factors together with genetic risk factors in the UK population are not known.

In order to inform future risk stratified screening approaches in the UK, in this study we assessed the performance of risk scores that include common genetic variants and predict future CRC, with or without phenotypic risk factors, in a cohort of 443,888 individuals.

The identification of genetic risk factors for common diseases has been disproportionately focused in white / European populations[15]. Consequently, there is concern that risk stratified screening in the UK could increase inequalities in health. We therefore also additionally assessed the performance of genetic risk scores among ethnic minorities in an exploratory analysis.

MATERIALS AND METHODS

We performed an external validation of genetic (GRS) and combined GRS plus phenotypic risk models for prediction of incident CRC, following the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guideline[16].

Validation cohort

For our validation cohort we used UK Biobank, the largest population-based cohort in the UK[17]. Women and men aged 37-73 years, who were registered with the National Health Service and lived within approximately 25 miles of one of 22 study assessment centres across the UK, were invited to participate between 2006-2011. From 9.2 million invitations 503,325 were recruited (5.5%) and attended a baseline assessment when data were collected using touchscreen questionnaires, interviews and physical measurements, and blood samples provided.

CRC diagnosis in the validation cohort

CRC cancer diagnosis is recorded for UK Biobank participants through linkage to national cancer registries. The most recent cancer record in UK Biobank has a diagnosis date of 27th October 2016. We censored all follow up at 31st March 2016 to ensure that late registrations were not missed. We included 451,171 out of 502,326 UK Biobank participants who had 6 years of complete follow up from date of baseline assessment until 31st March 2016, excluding people who had less than 6 years follow up but including those who died during the follow-up period. For this validation we identified CRC with the following diagnosis codes: ICD9 153.0-153.9, 154.0, 154.1 and 154.8 and ICD10 C18.0-C18.9, C19, C20 and C21.8. We excluded from all analyses 2,489 participants who had a diagnosis of CRC prior to the baseline assessment. 659 included participants had a CRC diagnosis after the end of the included 6 years of follow-up and are included in this analysis as non-cancer cases. As participants with previous colorectal polyps, n=1,473, or a diagnosis of inflammatory bowel disease (IBD), n=4,231, would likely be in surveillance programmes, we also excluded individuals with a self-reported history at baseline. Of the 443,888 participants included in our primary analysis there were 2,679 (0.6%) cases of incident CRC within the complete 6 year follow-up period.

The only exception to this was for the validation of one model (Huyghe et al.) [18] for which UK Biobank CRC cases from September 2014 and earlier formed part of the original analysis

data set for this risk prediction model. For this model we included all UK Biobank participants without a history of colorectal polyps or IBD who were still alive and without a CRC diagnosis on 30th September 2014, and followed them up until 31st March 2016. For this model we included 482,089 participants and 842 cases of CRC in the 18 month follow up period.

Selection of risk prediction models

In our systematic review[19] we identified 29 genetic only (GRS) or combined GRS and phenotypic models for CRC from 20 publications. After contacting authors if insufficient data were provided in the original publications, we excluded 11 models because either details of the risk alleles were not available[20,21], the relative risk parameters associated with predictors included in the model were not published[22–24], they included variables not available in UK Biobank[25–27] or biochemical risk factors[28], or if the model was developed separately for proximal and distal colon and rectal cancers[29].

After these exclusions, we included 17 models from our systematic review[18,22,24,30–39]. We additionally separately included the GRS component from five of these models[29–31,36,39] that were developed only as combined GRS plus phenotypic models. For the Hsu model[29], only the GRS was included as the GRS plus phenotypic versions were developed separately for proximal and distal colon, and rectal cancers. From the Smith publication we considered only the model incorporating the Wells phenotypic risk score[40]. We further included the GRS plus phenotypic risk model developed by Jenkins *et al.*, [41] that was published separately from the genes-alone GRS developed by the same team.

Details of these 23 models, including the study design, development method and the risk factors used in the prediction of CRC risk for each are given in Table 1. Fourteen models were GRSs including only SNPs, four included SNPs plus phenotypic factors, but not age, and five a combination of SNPs, phenotypic factors and age.

Genotyping

For all UK Biobank participants blood samples were genotyped using Affymetrix UK BiLEVE Axiom array and Affymetrix UK Biobank Axiom array and imputed to the combined 1000 Genomes Project v.3 and UK10K reference panels using SHAPEIT3 and

IMPUTE3. The lowest imputation info score for the SNPs used in these analyses was 0.86[42].

Coding of risk models in UK Biobank

Full details of the definition of each risk factor and how we operationalised them in the UK Biobank dataset and handled missing data are given in Supplementary Tables 1 (SNPs) and 2 (phenotypic risk factors).

Ethnicity

Race/ethnicity, hereafter ethnicity, as this is the terminology used in the UK, was quantified using categories from the UK Office of National Statistics 2001 16 group classification[43], summarised in 5 groups; *White/European*, (response options:

English/Welsh/Scottish/Northern Irish/British, Irish, and Any other White background), *Mixed* (White and Black Caribbean, White and Black African, White and Asian, Any other Mixed/multiple ethnic background), *South Asian* (Indian, Pakistani, Bangladeshi, Any other Asian background), *Black* (African, Caribbean, Any other Black/African/Caribbean background), *Other* (Chinese, Any other ethnic group).

Analysis

For all of the models, we computed the predicted risk score for developing CRC for each participant using data collected at baseline assessment. Only one score (Dunlop et al) predicted an absolute risk of CRC over a specific time frame, all other scores predicted relative, rather than absolute risk (RR).

We assessed model performance in terms of discrimination and relative risk calibration. For GRS scores we calculated standardized versions of the scores and estimated the odds ratio (OR) for CRC associated with a 1 standard deviation (SD) increase in the standardized risk scores as a measure of discrimination; discrimination was further assessed using the area under the curve (AUC) for all models.

Calibration was assessed graphically by comparing the expected with observed relative risk of developing CRC over the six-year follow-up period. We stratified our analysis cohort by

deciles of predicted relative risk (with decile 5 as the baseline). The observed relative risk in each decile was calculated as the proportion of cases in the decile divided by the proportion of cases in the entire sample, divided by the same estimate in decile 5. The expected RR in each decile was calculated as the geometric mean (because log RR rather than RR is normally distributed) of the predicted RR in the decile divided by the mean predicted RR in decile 5.

We also calculated sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-) and the positive and negative predictive values (PPV and NPV) using a cut-off value for each risk score chosen such that 10% of the population had values above the cut-off; the procedure was then repeated using cut-offs for which 20%, 80% and 90% had values above the cut-off.

In a further analysis we compared the performance of the GRS scores with models additionally including age (categorized as above and below age 60, the current UK CRC screening threshold age) and family history in order to provide a preliminary evaluation to inform whether there would be a benefit to incorporate GRS into clinical practice.

Our first analysis considering the discrimination of GRSs alone was stratified by ethnicity, and considered all UK Biobank participants without stratification by sex, because of the relatively small numbers of people reporting non-white/European ethnicity in UK Biobank. Discrimination using AUC and RR calibration were assessed among people from all ethnic backgrounds and stratified by sex. All analyses were carried out in Stata 15 (StataCorp 2017).

Missing data

For our primary analyses we used a “complete-case” approach including only those individuals for whom all of the risk factors in a particular prediction model were available. The sample size is therefore consistent across all genetic models, except for the model of Wang *et al* (that incorporated genotypes which, in a small number of samples, could not be ascertained from imputed allele counts), and the model of Huyghe *et al*. as described above, but sample sizes varied between models which included non-genetic risk factors, as the amount of missing data for phenotypic risk factors – particularly dietary ones – varied between models (Supplementary table 3).

Sensitivity analyses

We primarily focused our sensitivity analyses on areas not previously addressed[11] and those particularly relevant to genetic risk models. Accordingly, we compared the performance of risk scores after excluding people with high degree of relatedness to other cohort members (people with >10 relatives, estimated using genetic data), and randomly excluding one of each pair of first degree relatives. We also compared the performance in a cohort excluding all individuals with any cancer diagnosis prior to baseline. For both these analyses we followed the methods from previous work[44]. We carried out an additional sensitivity analysis comparing discrimination for the GRS models among people 60 and under, and over 60. In a further sensitivity analysis we excluded all people with a history of colonoscopy at baseline.

For GRS plus phenotypic models we carried out a further sensitivity analysis restricting our sample for the GRS models to those individuals in whom the GRS plus phenotypic version of the risk score could be estimated (i.e. only among those individuals with complete phenotype characterization for each score).

For the Huyghe model[18] we carried out a further sensitivity analysis including the 6 years of follow used for the other models. . For the model from Abe[30] we additionally compared results including and excluding folate, because of the large amount of missing data for this variable.

RESULTS

The characteristics of the study population are shown in Supplementary Table 4. Compared to those who did not develop CRC, those who did were more likely to be male, older, and have a family history of CRC. Although the number of SNPs varied across models, the median number of risk alleles among CRC cases was typically one higher than among those who did not develop cancer.

Discrimination

GRS discrimination, stratified by ethnicity

Among the 441,141 participants who self-reported their ethnicity, 419,579 (95%) were White. Mean standardized GRS among people with CRC varied between 0.02 and 0.45 in the whole validation cohort, and among the whole cohort varied from -1.44 to 0.85 when

stratified by ethnicity (Supplementary table 5). The Huyghe GRS had the highest OR per 1SD of GRS, 1.60 (1.50-1.72) among all UK Biobank participants and 1.60 (1.50-1.72) among people reporting a white/European ethnic background. Five of the GRSs showed some discriminative ability among people with non-white ethnicity[22,24,29,31,35] (Table 2).

GRS discrimination, stratified by sex

Figure 1 and Supplementary Table 6 show the discrimination, measured by AUC, for all models in the whole UK Biobank population, stratified by sex. Discrimination for the 14 GRS models is similar in men and women and ranges from 0.50 to 0.63. The AUC increased with the number of SNPs included within the risk models (Figure 2), with the 120-SNP model by Huyghe having the highest AUC of 0.62 (95% CI 0.59-0.64) in women and 0.64 (95% CI 0.61-0.66) in men. In general, the GRSs based on newer genome-wide association (GWA) studies (GWA) also performed better than those based on candidate genes or older GWA studies. These analyses stratified by White-European/non-White-European ethnicity are presented in Supplementary Table 7.

< insert figure 1 about here >

< insert figure 2 about here >

GRS-plus phenotype discrimination, stratified by sex

Among the four models that incorporated both genetic and phenotypic risk factors but not age[31,34,38,41], adding phenotypic factors, including BMI, smoking, family history, alcohol intake, red meat intake and physical activity, either did not change the AUC[31,34] or reduced it[38,41] in women. In men, however, the AUC improved from 0.55 (0.54-0.57) to 0.59 (0.57-0.6) with the addition of smoking, BMI, alcohol, fibre intake, red meat intake and physical activity in the Yarnall model[34] and from 0.55 (0.54-0.57) to 0.58 (0.56-0.59) with the addition of family history, BMI, alcohol intake, physical exercise, red meat and vegetable intake and NSAIDs/aspirin use in the Ibanez-Sanz model[38].

Considering only people in whom it was possible to calculate a risk-score based on both the GRS and GRS plus phenotypic versions of the same models confirmed these differences, with significant ($p<0.05$) improvements in AUC observed only with the addition of phenotypic risk factors in the Yarnall and Ibanez-Sanz models in men (Supplementary Table 8).

By comparison, the discrimination of models including age was greater ($p<0.05$) than for GRSs alone for all models in both women and men, with a relatively greater improvement in performance in men than women (Figure 1). The models with highest AUC were those developed by Abe, Hosono, Dunlop and Smith[24,30,32,36], with AUCs between 0.67-0.71 in men and 0.64-0.67 in women.

The sensitivity, likelihood ratios and PPV and NPV for thresholds at which varying percentages of the population are classified as high risk are presented in full in Supplementary Tables 9 and 10. Within the 10% of the population with the highest risk, the top performing GRS[18] identified 18.6% of women and 22.3% of men who went on to develop CRC. Among those with the highest 20% risk, this increased to 30.7% for women and 35.7% for men. For the four best performing GRS plus phenotypic models that also included age[24,30,32,36], in women the top 10% of the population included between 19.7% and 20.7% of those who went on to develop CRC and the top 20% 33.8% to 38.1%. In men, the corresponding values are 21.1% to 26.9% and 38.2% to 47.9%. The NPVs were high and comparable (>99.3) for all models.

Calibration

Results from the relative risk calibration are presented for women and men in Figures 3 and 4 and Supplementary Tables 11 and 12. There is variation in calibration across models, with some models having very poor calibration. There is a consistent pattern that models tend to over-estimate relative risk at higher levels of risk. Calibration did not substantially improve in models based on a GRS plus phenotypic risk factors compared with those that include only a GRS, except for those which incorporated age.

< insert figure 3 about here >

< insert figure 4 about here >

The additional analysis comparing the AUC of models incorporating each GRS, age and family history, with age and family history alone found that for all models except Wang adding the GRS increased the AUC by between 0.01 and 0.03 (Supplementary Table 13), i.e. reducing between 3-9% of the remaining error.

Sensitivity analyses

The results from all the other sensitivity analyses were consistent with the main analysis (Supplementary Tables 14-19).

DISCUSSION

Key findings

This is, to our knowledge, the first external validation to use a single cohort to directly compare multiple published risk prediction models for CRC that include common genetic markers. It shows that genetic information alone discriminates moderately well between those who do and do not develop CRC over a six year period. The best performing GRS by Huyghe *et al.*, [18] had an AUC of 0.62 in women and 0.64 in men and OR per 1SD of 1.60, comparable with polygenic risk scores in breast cancer (AUC 0.63 (95% CI 0.62-0.64), OR per 1SD 1.49-1.71 [45]) and coronary artery disease (0.62 (95% CI: 0.62 to 0.63) [46]) and better than risk models including phenotypic risk factors without age [11,24].

In contrast to risk models incorporating phenotypic risk markers for which performance in men is better than in women [11,12], the performance of GRSs was equivalent. Consistent with this, adding phenotypic risk factors without age to a GRS improved the AUC in men but not in women, with the best performing models including SNPs, phenotypic risk factors and age having AUCs between 0.67-0.71 in men and 0.64-0.67 in women.

The potential impacts of incorporating these risk scores into practice are best appreciated by the differences in risk classification between the models. Our previous work using UK Biobank identified that using the current age-based English bowel cancer screening threshold of 60 years to identify the population at high risk for CRC and then random sampling to

account for ties, targeting people in the top 10% of risk would identify 17% of men and 16% of women who developed CRC[11]. The equivalent proportions using the Huyghe model with the highest discrimination for genetic risk factors alone in this study would be 22% for men and 19% for women, representing an increase in detection of 2-5% of cases for the same number of people screened. This is consistent with an analysis by Jenkins et al.,[13] who showed that, in a hypothetical population based on the Australian population in 2011, inviting individuals for screening based on their genetic risk rather than all those aged between 50 and 74 would result in a 3.1% overall improved efficiency.

Strengths and Limitations

The main strengths of this analysis are the use of a large cohort with nearly 3 million person-years of follow-up, comprehensive phenotyping and genotyping and linkage to national cancer registries, and the inclusion of 23 risk models identified from a systematic review[19]. We were not only able to perform the largest external validation to date of multiple CRC risk models in a single population but were also able to compare the performance of models including only SNPs to those including SNPs plus phenotypic risk factors with or without age and family history.

However, despite its many strengths, the UK Biobank, has a number of limitations. In particular, the response rate to invitations to take part was only 5.5%[17]. As a result, participants are more likely to be older, to be female, and to live in less socioeconomically deprived areas than non-participants; incidence rates of CRC are also lower than in the general population[47]. We limited the effect of this ‘healthy volunteer’ bias by restricting our analyses to relative risk and discrimination. Analysis of SNP data may also potentially be less likely to be confounded by this ‘healthy volunteer’ bias than models including phenotypic risk factors as well. The large variation in age at baseline assessment also means that age has a disproportionately large impact on CRC risk compared other phenotypic measures. A third limitation is that CRC cases registered at Scottish cancer registries may not all be complete after 31/03/2015. This may impact up to 10% of the cohort and result in an underestimate of CRC risk overall. Finally, CRC screening at age 60 in the UK was rolled out during the period of recruitment to UK Biobank. Some, but not all members of the cohort will therefore have participated in the national screening programme during this period. Our

sensitivity analysis excluding people with a history of colonoscopy at baseline, however, did not change the findings of this analysis.

Although similar to the 2011 UK Census[47], the predominantly White population within the UK Biobank cohort also means that AUC estimates when restricted to participants from ethnic minorities have wide confidence intervals and assessing the model performance is challenging. Any small differences between White and non-White populations should, therefore, not be over-interpreted. In our primary analysis we have applied the GRS/GRS plus phenotypic risk models to everyone, regardless of the population in which they were developed. We acknowledge the issue of population stratification leading to spurious associations between genetic risk factors and disease when people from diverse ethnicities are analysed in the same cohort. Nonetheless, in practice any policy for developing stratified screening will be applied across the whole country; these results present the performance of these models in this scenario.

There are also limitations with our analysis. Firstly, we excluded two models identified from our systematic review because they included variables not present in UK Biobank and we had to derive proxy variables if there were no exact matches for variables included in the risk models. We also only assessed RR calibration. Although this does not allow us to assess how closely the predicted absolute risk matches the observed absolute risk, it illustrates the variation in calibration across the models and highlights that models would need to be calibrated to the population in which they are going to be used.

Comparison with existing literature

When compared with findings of validation studies in external populations or non-random split samples[19], our results for discrimination are similar for the genetic risk models by Dunlop, Ibanez-Sanz and Smith, but lower for the genetic risk models by Hosono and Xin. The lower discrimination seen in this cohort for the Hosono and Xin models likely reflects differences in the ethnicity of the populations: the model by Hosono was developed using logistic regression and subsequently validated in a Japanese cohort and the model by Xin used published GWAS studies from European and Asian populations and validated the model in a Chinese cohort. Our results for the Smith model that combined genetic risk factors with phenotypic risk factors and age from the model by Wells were also consistent with the evaluation of that model in an earlier release of UK Biobank data. For the remaining three

models (Abe, Dunlop and Hosono) our results showed better discrimination. In all three cases the previous validation studies had been performed in case-control studies matched by age, effectively removing the effect of age from the models, and highlighting the strength of age as a risk factor. This effect of age may also explain why Smith *et al.* found in UK Biobank that addition of a GRS to either the models by Taylor[48] or Wells[40], both of which include age, did not improve discrimination and did not result in a substantive change in the predicted probability for the majority of participants[24].

Conclusions and implications for future research

This study shows that existing GRSs are able to discriminate moderately well between those who develop CRC and those that do not and, unlike phenotypic risk scores, they do not perform differentially in men and women. Genetic risk scores may also be easier to implement than phenotypic risk scores as they do not change over time and are not associated with the same degree of measurement error and bias as phenotypic risk factors. With ongoing GWAS studies and increasingly comprehensive imputation panels that allow for improved low-frequency and rare genetic variant imputation, it is expected that new SNPs associated with CRC will be identified in the future[35,49]. Alongside these efforts to identify further SNPs, future studies are needed to develop and validate GRSs in non-European populations and model the potential impact of incorporating them into screening programmes, as is being done for other cancers[8]. There are a number of other issues that also need to be considered before stratified screening based on genetic risk can be implemented. These include the need for fundamental changes in the infrastructure and mechanisms for genetic data collection, storage and sharing[50], and ethical, legal and social considerations such as equity of access to genetic testing, insurance issues, and whether screening programmes should exclude people currently eligible as a result of their age on the basis of low genetic risk. Further research should focus on these areas as well as modelling the potential health benefits and cost effectiveness of implementing stratified genetic risk based colorectal cancer screening. By demonstrating the current performance of risk scores including genetic risk information in a large UK population, this study supports the need for such research.

ACKNOWLEDGEMENTS

We thank James Brimicombe for data management support and our patient and public representative Margaret Johnson for providing helpful comments on the findings.

Ethical approval

The UK Biobank study was approved by the North West Multi- Centre Research Ethics Committee (reference number 06/MRE09/ 65), and at recruitment all participants gave informed written consent to participate in UK Biobank and be followed up, using a signature capture device.

Data sharing

All data on the risk models are available from the reports or authors of the primary research. All key data fields derived for this study and the underlying STATA code used to generate the main results of the paper will be made available through the UK Biobank access team (access@ukbiobank.ac.uk). No additional data are available.

REFERENCES

- 1 Gunter MJ, Alhomoud S, Arnold M, *et al.* Meeting report from the joint IARC–NCI international cancer seminar series: a focus on colorectal cancer. *Ann Oncol* 2019;**30**:510–9. doi:10.1093/annonc/mdz044
- 2 Gutlic I, Schyman T, Lydrup M-L, *et al.* Increasing colorectal cancer incidence in individuals aged < 50 years—a population-based study. *Int J Colorectal Dis* 2019;;1221–6. doi:10.1007/s00384-019-03312-3
- 3 Lin JS, Piper MA, Perdue LA, *et al.* Screening for Colorectal Cancer. An Updated Systematic Review for the U.S. Preventive Services Task Force. *JAMA* 2016;**315**:2576. doi:10.1001/jama.2016.3332
- 4 Garborg K, Holme Ø, Løberg M, *et al.* Current status of screening for colorectal cancer. *Ann Oncol* 2013;**24**:1963–72. doi:10.1093/annonc/mdt157
- 5 Kronborg O, Fenger C, Olsen J, *et al.* Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996;**348**:1467–71. doi:10.1016/S0140-6736(96)03430-7
- 6 Lindholm E, Brevinge H, Haglund E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. *Br J Surg* 2008;**95**:1029–36. doi:10.1002/bjs.6136
- 7 Schreuders EH, Ruco a., Rabeneck L, *et al.* Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015;;1637–49. doi:10.1136/gutjnl-2014-309086
- 8 Pashayan N, Morris S, Gilbert FJ, *et al.* Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA Oncol* Published Online First: 2018. doi:10.1001/JAMAONCOL.2018.1901
- 9 Tammemägi MC. Selecting lung cancer screenees using risk prediction models—where do we go from here. *Transl Lung Cancer Res* 2018;**7**:243–53. doi:10.21037/tlcr.2018.06.03
- 10 Meisel SF, Side L, Fraser L, *et al.* Population-based, risk-stratified genetic testing for ovarian cancer risk: A focus group study. *Public Health Genomics* 2013;**16**:184–91. doi:10.1159/000352028
- 11 Usher-Smith J, Harshfield A, Saunders C, *et al.* External validation of risk prediction models for incident colorectal cancer using UK Biobank. *Br J Cancer* 2018;;1–10. doi:10.1038/bjc.2017.463
- 12 Smith T, Muller DC, Moons KGM, *et al.* Comparison of prognostic models to predict the occurrence of colorectal cancer in asymptomatic individuals: A systematic literature review and external validation in the EPIC and UK Biobank prospective cohort studies. *Gut* 2019;**68**:672–83. doi:10.1136/gutjnl-2017-315730
- 13 Stanesby O, Jenkins M. Comparison of the efficiency of colorectal cancer screening programs based on age and genetic risk for reduction of colorectal cancer mortality. *Eur J Hum Genet*

- 2017;**25**:832–8. doi:10.1038/ejhg.2017.60
- 14 Frampton MJE, Law P, Litchfield K, *et al.* Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol* 2016;**27**:429–34. doi:10.1093/annonc/mdv540
- 15 Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell* 2019;**177**:26–31. doi:10.1016/j.cell.2019.02.048
- 16 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;**162**:55–63. doi:10.7326/M14-0697
- 17 Allen N, Sudlow C, Downey P, *et al.* UK Biobank: Current status and what it means for epidemiology. *Heal Policy Technol* 2012;**1**:123–6. doi:10.1016/j.hlpt.2012.07.003
- 18 Huyghe JR, Bien SA, Harrison TA, *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;**51**:76–87. doi:10.1038/s41588-018-0286-6
- 19 McGeoch L, Saunders CL, Griffin SJ, *et al.* Risk prediction models for colorectal cancer incorporating common genetic variants: a systematic review. *Cancer Epidemiol Biomarkers Prev* Published Online First: 10 July 2019. doi:10.1158/1055-9965.EPI-19-0059
- 20 Li J, Chang J, Zhu Y, *et al.* Risk prediction of colorectal cancer with common genetic variants and conventional non-genetic factors in a Chinese Han population. *Zhonghua Liu Xing Bing Xue Za Zhi* 2015;**36**:1053–7.
- 21 Jo J, Nam CM, Sull JW, *et al.* Prediction of Colorectal Cancer Risk Using a Genetic Risk Score: The Korean Cancer Prevention Study-II (KCPS-II). *Genomics Inform* 2012;**10**:175–83. doi:10.5808/GI.2012.10.3.175
- 22 Xin J, Chu H, Ben S, *et al.* Evaluating the effect of multiple genetic risk score models on colorectal cancer risk prediction. *Gene* 2018;**673**:174–80. doi:https://dx.doi.org/10.1016/j.gene.2018.06.035
- 23 Jung S, Yon PhD MPH, Zhang Z-F, *et al.* The effects of genetic variants related to insulin metabolism pathways and the interactions with lifestyles on colorectal cancer risk. *Menopause* Published Online First: 2019. doi:10.1097/GME.0000000000001301
- 24 Smith T, Gunter MJ, Tzoulaki I, *et al.* The added value of genetic information in colorectal cancer risk prediction models: development and evaluation in the UK Biobank prospective cohort study. *Br J Cancer* Published Online First: 2018. doi:10.1038/s41416-018-0282-8
- 25 Weigl K, Thomsen H, Balavarca Y, *et al.* Genetic Risk Score Is Associated With Prevalence of Advanced Neoplasms in a Colorectal Cancer Screening Population. *Gastroenterology* 2018;**155**:88–98.e10. doi:https://dx.doi.org/10.1053/j.gastro.2018.03.030
- 26 Procopciuc LM, Osian G, Iancu M. Colorectal Cancer Carcinogenesis: a Multivariate Genetic Model in a Cohort of Romanian Population. *Clin Lab* 2017;**63**:647–58. doi:https://dx.doi.org/10.7754/Clin.Lab.2016.160821
- 27 Shiao SPK, Grayson J, Lie A, *et al.* Personalized nutrition—genes, diet, and related interactive

- parameters as predictors of cancer in multiethnic colorectal cancer families. *Nutrients* 2018;**10**:1–19. doi:10.3390/nu10060795
- 28 Jung KJ, Won D, Jeon C, *et al.* A colorectal cancer prediction model using traditional and genetic risk scores in Koreans. *BMC Genet* 2015;**16**:49. doi:http://dx.doi.org/10.1186/s12863-015-0207-y
 - 29 Hsu L, Jeon J, Brenner H, *et al.* A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology* 2015;**148**:1330. doi:https://dx.doi.org/10.1053/j.gastro.2015.02.010
 - 30 Abe M, Ito H, Oze I, *et al.* The more from East-Asian, the better: risk prediction of colorectal cancer risk by GWAS-identified SNPs among Japanese. *J Cancer Res Clin Oncol* 2017;**143**:2481–92. doi:10.1007/s00432-017-2505-4
 - 31 Jeon J, Du M, Schoen RE, *et al.* Determining Risk of Colorectal Cancer and Starting Age of Screening Based on Lifestyle, Environmental, and Genetic Factors. *Gastroenterology* 2018;**154**:2152–2164.e19. doi:10.1053/j.gastro.2018.02.021
 - 32 Dunlop MG, Tenesa A, Farrington SM, *et al.* Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut* 2013;**62**:871–81. doi:10.1136/gutjnl-2011-300537
 - 33 Wang HM, Chang TH, Lin FM, *et al.* A new method for post Genome-Wide Association Study (GWAS) analysis of colorectal cancer in Taiwan. *Gene* 2013;**518**:107–13. doi:10.1016/j.gene.2012.11.067
 - 34 Yarnall JM, Crouch DJM, Lewis CM. Incorporating non-genetic risk factors and behavioural modifications into risk prediction models for colorectal cancer. *Cancer Epidemiol* 2013;**37**:324–9. doi:10.1016/j.canep.2012.12.008
 - 35 Frampton MJE, Law P, Litchfield K, *et al.* Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol* 2016;**27**:429–34. doi:http://dx.doi.org/10.1093/annonc/mdv540
 - 36 Hosono S, Ito H, Oze I, *et al.* A risk prediction model for colorectal cancer using genome-wide association study-identified polymorphisms and established risk factors among Japanese. *Eur J Cancer Prev* 2016;**25**:500–7. doi:10.1097/CEJ.0000000000000213
 - 37 Jenkins M., Makalic E, Dowty J., *et al.* Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening. *Futur Oncol* 2016;**12**:503–13. doi:http://dx.doi.org/10.2217/fon.15.303
 - 38 Ibanez-Sanz G, Diez-Villanueva A, Alonso MH, *et al.* Risk Model for Colorectal Cancer in Spanish Population Using Environmental and Genetic Factors: Results from the MCC-Spain study. *Sci Rep* 2017;**7**:43263. doi:https://dx.doi.org/10.1038/srep46807
 - 39 Iwasaki M, Tanaka-Mizuno S, Kuchiba A, *et al.* Inclusion of a genetic risk score into a validated risk prediction model for colorectal cancer in Japanese men improves performance.

- Cancer Prev Res* 2017;**10**:535–41. doi:<https://dx.doi.org/10.1158/1940-6207.CAPR-17-0141>
- 40 Wells BJ, Kattan MW, Cooper GS, *et al.* ColoRectal Cancer Predicted Risk Online (CRC-PRO) Calculator Using Data from the Multi-Ethnic Cohort Study. *J Am Board Fam Med* 2014;**27**:42–55. doi:10.3122/jabfm.2014.01.130040
- 41 Jenkins MA, Win AK, Dowty JG, *et al.* Ability of known susceptibility SNPs to predict colorectal cancer risk for persons with and without a family history. *Fam Cancer* Published Online First: 2019. doi:10.1007/s10689-019-00136-6
- 42 Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;**562**:203–9. doi:10.1038/s41586-018-0579-z
- 43 Office for National Statistics. Ethnic group, national identity and religion. <https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion> (accessed 7 Jan 2020).
- 44 Mavaddat N, Michailidou K, Dennis J, *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019;**104**:21–34. doi:10.1016/j.ajhg.2018.11.002
- 45 Mavaddat N, Michailidou K, Dennis J, *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2018;:1–14. doi:10.1016/j.ajhg.2018.11.002
- 46 Inouye M, Abraham G, Nelson CP, *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults. *J Am Coll Cardiol* 2018;**72**:1883–93. doi:10.1016/j.jacc.2018.07.079
- 47 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* 2017;**186**:1–9. doi:10.1093/aje/kwx246
- 48 Taylor DP, Stoddard GJ, Burt RW, *et al.* How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. *Genet Med* 2011;**13**:385–91. doi:10.1097/GIM.0b013e3182064384
- 49 Zhang Y, Wilcox AN, Zhang H, *et al.* Assessment of Polygenic Architecture and Risk Prediction based on Common Variants Across Fourteen Cancers. *bioRxiv* 2019;:723825. doi:10.1101/723825
- 50 Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. *Nature* 2015;**526**:336–42. doi:10.1038/nature15816

Table 1. Details of the development and factors included in each of the risk scores included in validation study

Author, year	Country	Genetic risk score			Genetic plus phenotypic risk score						
		Number of SNPs	Method of development of GRS	Developed in original publication as a "genes-only" score	Method of development of combined model	Non-genetic risk factors included in score					
						Age	Sex	FH	BMI	Smoking	Other
Abe 2017	Japan (d, v)	11	Unweighted allele counting model		Logistic regression	•	•	•	•	•	Referral pattern, alcohol consumption, regular exercise, and dietary folate intake
Dunlop 2013	UK, Canada, Australia, USA and Germany (d) Sweden and Finland (v)	10	Weighted allele model weighted by published log odds	•	Stratified absolute risks calculated using logistic regression applied to Scottish data	•	•	•			
Frampton 2016	UK (v)	37	Weighted allele model weighted by published log odds	•							
Hosono 2016	Japan (d, v)	6	Unweighted allele counting model		Logistic regression	•		•	•	•	Referral pattern, alcohol consumption, regular exercise, dietary folate intake
Hsu 2015	USA and Germany (d, v)	31	Weighted model weighted by published log-odds								
Huyghe 2019	European (91.7%) and East Asian (8.3%) (d)	120	Weighted allele model weighted by study derived weights	•							
Ibanez-Sanz 2017	Spain (d, v)	21	Unweighted allele counting model (weighted allele models weighted by published log-odds and study derived log-odds similar so not reported)	•	Logistic regression			•	•		Alcohol use, physical exercise, red meat and vegetable intake, NSAIDs/aspirin use
Iwasaki 2017	Japan (d, v)	6	Weighted allele model weighted by study derived log-transformed per allele HR		Weighted cox proportional hazards regression	•	Developed in men only		•	•	Alcohol

Jenkins 2016 / 2019	Weighted allele model weighted by published log odds	49	Weighted allele model weighted by published log odds	•				•			
Jeon 2018	Australia, Canada, Germany, Israel and USA.(d, v)	63	Weighted allele model weighted by study derived estimated regression coefficients		Logistic regression		Developed separately in women and men		•	•	Height, education, history of type 2 diabetes mellitus, alcohol consumption, regular aspirin use, regular NSAID use, smoking, intake of fibre, calcium, folate, processed meat, red meat, fruit, vegetables, total-energy, physical activity (both) HRT (women)
Smith 2018	UK (d, v)	42	Weighted allele model weighted by published log odds	•	Log GRS combined with predicted log hazard ratio from original model	•		•	•	•	Diabetes, multi-vitamin usage, years of education, alcohol intake, physical activity, NSAID usage, red meat intake, smoking, oestrogen use (women only)
Wang 2013	Taiwan (d, v)	16	Logistic regression / GRS based on genotypes not alleles	•							
Xin 2018	China (d, v)	14	Weighted allele models weighted by published log odds and by study derived weights	•							
Yarnall 2013	UK (v)	14	Weighted allele model / developed with a simulation based procedure using REGENT software	•	OR estimated from simulation based procedure using REGENT software				•	•	Alcohol, fibre intake, red meat intake, physical activity

CRC – colorectal cancer, SNP - single-nucleotide polymorphism, BMI – body mass index, NSAID – non-steroidal anti-inflammatory drug, GRS – genetic risk score. d = development; v – validation;

Table 2. Genetic Risk Score (GRS) discrimination / odds ratio for colorectal cancer per standard deviation increase in risk score, stratified by self-reported white / non-white ethnicity

	Mean standardised GRS (SD), n=430,511		OR (95%CI) per 1SD of GRS		
	Without CRC (n=427,908)	With CRC (n=2,603)	Only self-reported white ethnicity (n=419,579)	Only non-white ethnic backgrounds (n=21,562)	All (n=430,511)
Abe 2017	-0.01 (1.00)	0.16 (0.98)	1.20 (1.15 - 1.25)	1.23 (0.98 - 1.53)	1.18 (1.14 - 1.22)
Dunlop 2013	0.00 (1.00)	0.22 (0.98)	1.25 (1.20 - 1.30)	1.22 (0.98 - 1.52)	1.25 (1.20 - 1.30)
Frampton 2016	0.00 (1.00)	0.17 (1.00)	1.18 (1.13 - 1.23)	1.40 (1.12 - 1.76)	1.18 (1.14 - 1.23)
Hosono 2016	0.00 (1.00)	0.13 (0.97)	1.16 (1.12 - 1.21)	1.10 (0.88 - 1.36)	1.14 (1.10 - 1.19)
Hsu 2015	-0.01 (1.00)	0.26 (0.98)	1.30 (1.25 - 1.35)	1.27 (1.01 - 1.61)	1.30 (1.25 - 1.35)
Huyghe 2019	-0.01 (1.00)	0.47 (0.99)	1.60 (1.50 - 1.72)	1.10 (0.73 - 1.65)	1.60 (1.50 - 1.72)
Ibanez-Sanz 2017	0.00 (1.00)	0.19 (1.00)	1.20 (1.16 - 1.25)	1.28 (1.02 - 1.60)	1.22 (1.17 - 1.26)
Iwasaki 2017	0.00 (1.00)	0.13 (1.00)	1.15 (1.11 - 1.20)	0.96 (0.75 - 1.21)	1.14 (1.10 - 1.19)
Jenkins 2016	-0.01 (1.00)	0.23 (0.97)	1.28 (1.23 - 1.33)	1.10 (0.88 - 1.38)	1.26 (1.21 - 1.31)
Jeon 2018	0.00 (1.00)	0.29 (0.99)	1.33 (1.28 - 1.38)	1.42 (1.13 - 1.79)	1.33 (1.28 - 1.39)
Smith 2018	-0.01 (1.00)	0.23 (0.98)	1.27 (1.22 - 1.32)	1.39 (1.11 - 1.73)	1.26 (1.22 - 1.31)
Wang 2013	0.00 (1.00)	0.02 (0.98)	1.01 (0.97 - 1.05)	1.08 (0.85 - 1.36)	1.02 (0.98 - 1.06)
Xin 2018	0.00 (1.00)	0.13 (1.00)	1.15 (1.11 - 1.20)	1.24 (1.04 - 1.49)	1.14 (1.09 - 1.18)
Yarnall 2013	0.00 (1.00)	0.19 (1.01)	1.21 (1.17 - 1.26)	1.22 (0.98 - 1.53)	1.22 (1.17 - 1.26)

TITLES AND LEGENDS TO FIGURES

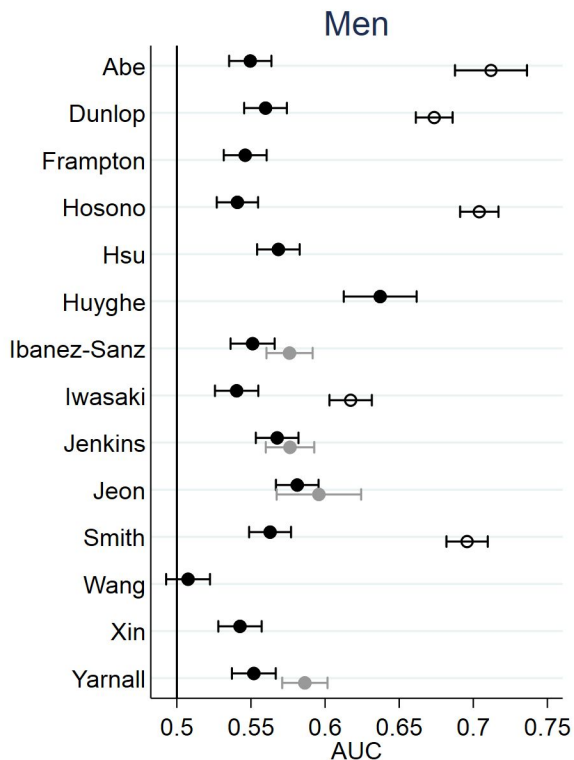
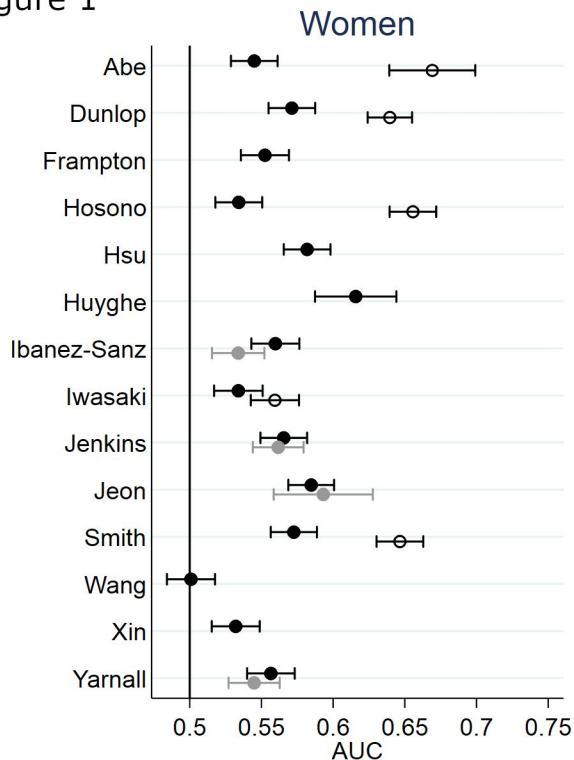
Figure 1. Discrimination, measured by the area under the receiver operating characteristic curve (AUC) for the risk models in women and men.

Figure 2. Discrimination ($AUC \pm 95\% \text{ CI}$) of models including genes only plotted against number of SNPs included in each model

Figure 3. Relative risk (RR) calibration in women.

Figure 4. Relative risk (RR) calibration in men.

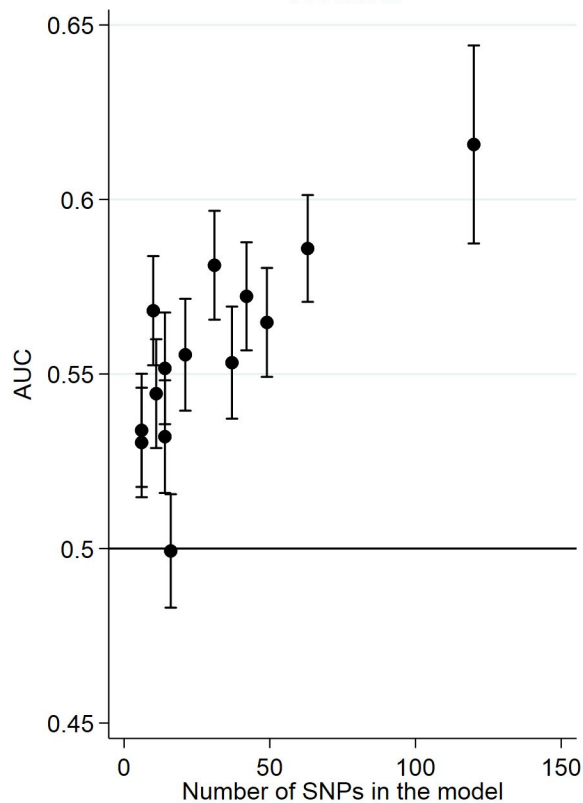
Figure 1



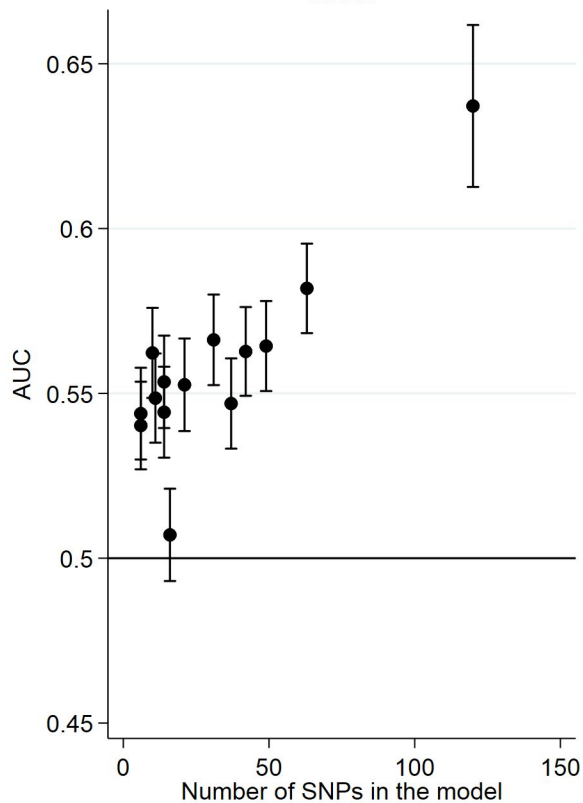
● Genes-only ○ Genes plus phenotypes (including age) ● Genes plus phenotypes (not including age)

Figure 2

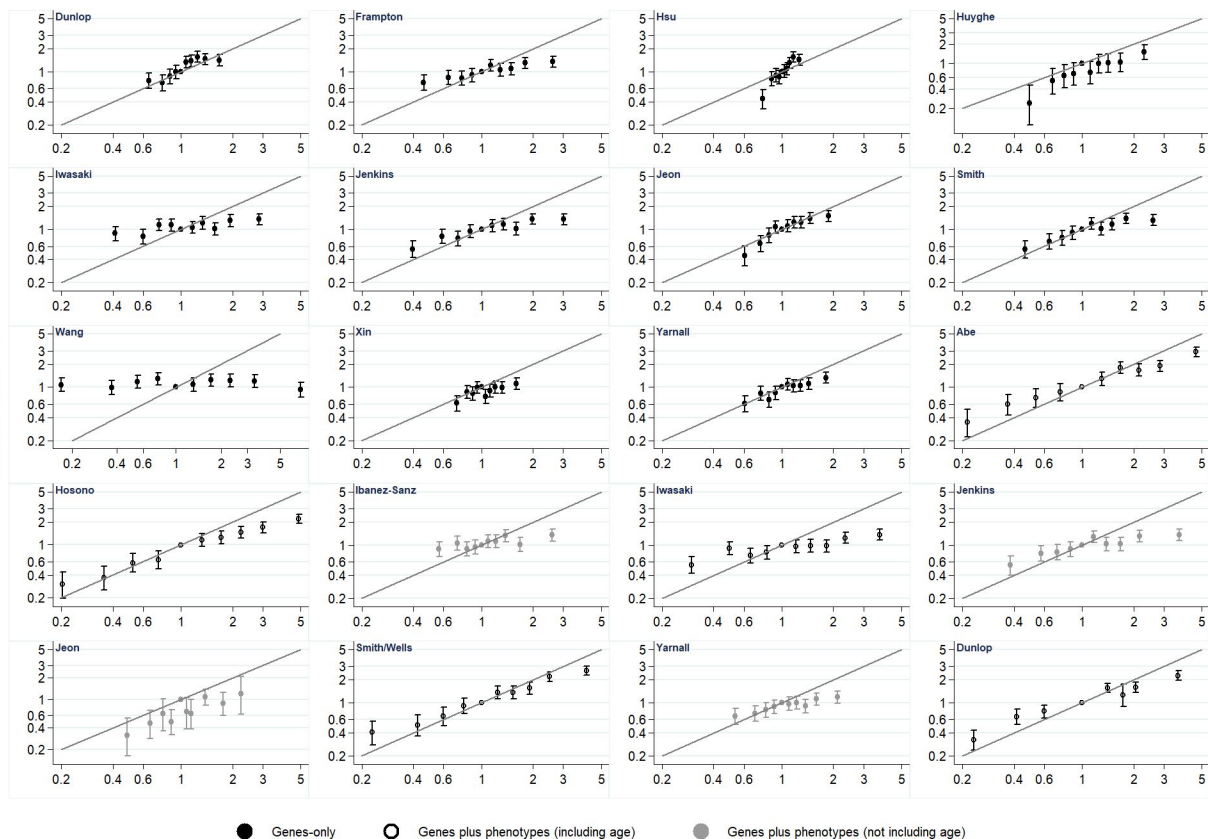
Women



Men



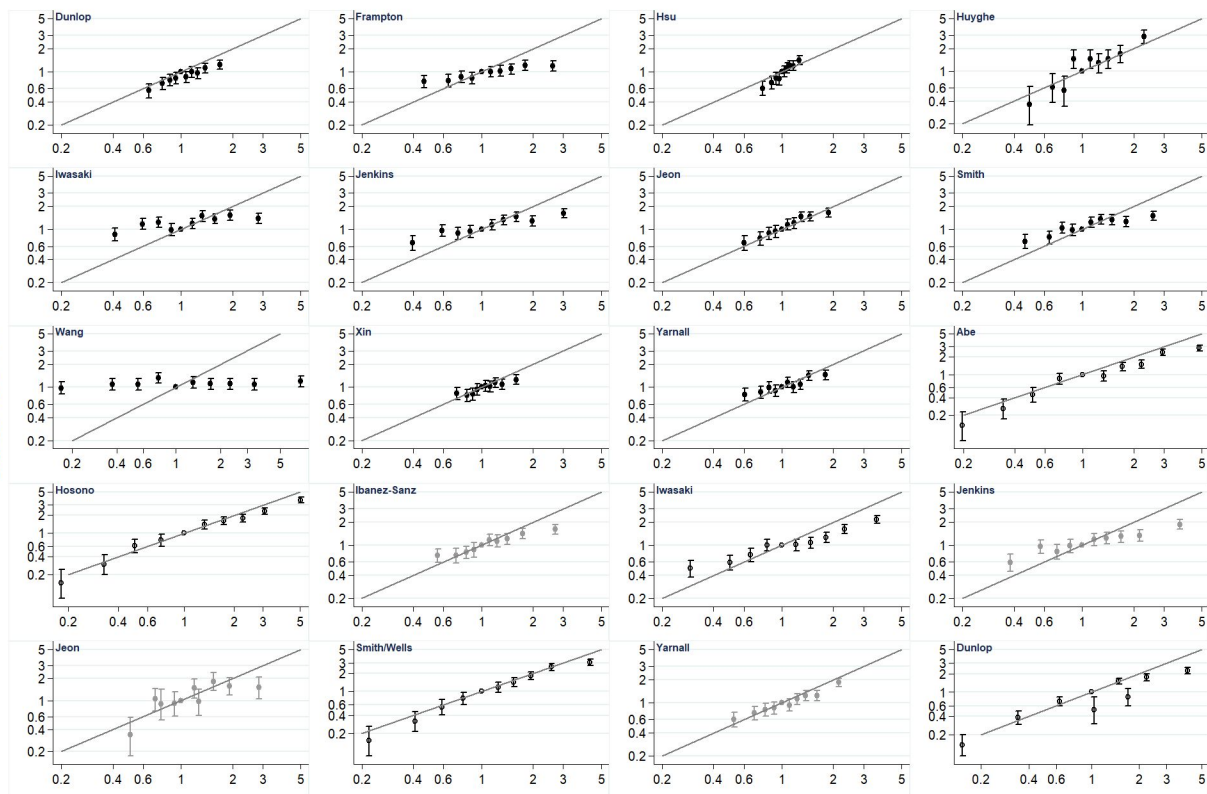
Observed RR



Expected RR

Figure 3

Observed RR



Genes-only



Genes plus phenotypes (including age)



Genes plus phenotypes (not including age)

Expected RR

Figure 4